LUND UNIVERSITY
Humanities and Theology

Centre for languages and literature

"Anyway, staraus ne sidet doma."

An explorative study of the Russian language in sms

Richard Hylerstedt

<u>Abstract</u>

This thesis aims to explore how the Russian language is used in text messaging on mobile phones. The focus of my work is strictly linguistic, and does not involve analysis of content or context. Within the linguistic area I have studied various aspects, for example orthography, reductions and style. I have gathered a corpus database consisting of 272 sms, donated by a group of Russians (mostly students in their early 20s). I am using a combination of qualitative and quantitative methodology to analyse the corpus and illustrate my findings. Brevity, symbols, omissions, non-standard spelling and informality are all features of sms language found in previous studies, and I have found these to be present in my corpus as well. The clearest tendency I have observed from my corpus is that Russian displays the same characteristics as other languages, but not to the same extent. I have found that Russians do not stray far from proper orthography in their messages and that some shortening strategies, such as consonant writing, are virtually inexistent. Another key feature, when looking at the corpus, is the widespread use of the Latin alphabet. People write their messages with Latin letters using different systems of transliteration, with a varying degree of 'officialness' and consequence. Possible reasons why Russian messages are more conservative include the relatively free structure of the language and the lesser exposure of Russian people to Western culture.

Table of contents

Introduction – No sportzal for me zavtra

"Ай донт ноу... Ай вас веры сурпрайсед майселф :)"

Most people would agree that sending a text message like the one above is a very complicated way to tell someone that you are confused by something unexpected ("I don't know… I was very surprised myself :)"), yet literally millions of messages, which are in principle similar to this one, are sent and received daily by Russians who use the Latin alphabet on their mobile phones.

It is difficult for native users of the Latin alphabet to imagine what it is like to write your mother tongue with a foreign alphabet, where many sounds of the language cannot be properly expressed. Inconvenient and sometimes imprecise compromises have to be made and each person will make up their own ad hoc solutions to transliteration issues. There are simply no international regulations for how to ask when dinner will be ready.

Most aspects of computer mediated communication has been fairly well adapted to the Cyrillic world, but due to certain circumstances the use of Latin letters in Russian sms communication is still widespread. Add to this the specific characteristics of interaction via text messages and the result is a unique application and adaptation of the Russian language. A niche that has not yet received the examination it deserves. The issue is of interest because of its unique features and because sms represent a frontier of language.

It is my aim and purpose with this study to explore the use of Russian in sms. I have imposed certain restrictions and reached certain limitations in my work but my overall research question is nonetheless the basic: How is Russian used in sms? Considering the depth of my investigation it is also my hope and intention that it will serve as an overview of the field and inspire further and deeper inquiries.

A case study of Russian sms language

My ambition is to present a case study of Russian sms language with a clear focus on fundamental linguistic properties. I do not wish to venture into the field of sociolinguistics, which is otherwise quite common in studies of sms. Neither do I seek to perform any content analysis on my messages or draw any conclusions based on the geographical location or gender of my informants. In short, I am not interested in what is said, when, why, by who or how often, but simply in how it is said. Within the area of language usage, I am studying the general dimensions associated with sms as well as the particularities of reductions and orthography.

To ensure some width in my narrowly defined field, I am looking at all types of sms and all stylistic levels. Apart from the general direction stated above, I have not set any restrictions as to what specific linguistic features to investigate.

The keywords when describing my methods are: explorative, empirical, descriptive and comparative. Overall, I have applied a rather open approach to the research process not to run the risk of obscuring any possible results by my choice of methods and perspectives. I am not setting out to prove a specific assumption, but rather to do a broad investigation and present what I find most interesting. At the beginning of my study I had no far-reaching expectations of what to find in the end. My work has been empirical by necessity and descriptive in the sense that I am not attempting to decide whether one way of using the language is better than the other. The comparative element of my study lies not only in contrasting different linguistic aspects with each other but also in the question about how Russian is used in sms compared to other languages. I find myself in the borderland between qualitative and quantitative research methodology in the sense that I am looking very closely at a relatively small set of samples but at same time trying to make some comparison of what features are more common.

The previous research done on language in sms is not as extensive as one could imagine. Further, there is a great deal of knowledge that has not yet been "academically solidified". Many accounts regarding sms are more popular, essayistic or even commercial in their approach. I have gathered material from

both printed and electronic sources, ranging form encyclopaedias to conference presentations, to produce the necessary background for my study.

How I collected and analysed 272 Russian sms

My guiding principle in selecting the informants for my study was to find people from different parts of Russia who represented a similar age group and educational background. I also wanted this group to consist of both men and women. By contacting people via my personal network of friends and via an online community for intercultural exchange I managed to recruit nine people who met these criteria satisfactorily. All informants are within the age span of 20 to 26, with the exception of one (aged 36), and a majority of them are university students. The ratio between men and women is 4 to 5 and the same holds true for residents of bigger cities (Moscow and Saint Petersburg) and more remote areas, such as Vladivostok and Kaliningrad. The average member of this group sends approximately three sms per day.

Finding people willing to share their messages was a lesser problem than finding those who had something to share, as sms proved to be a less common mean of communication among Russian students than expected. Another problem related to the selection of informants is of course that the results of and conclusions drawn from my study will only be valid for this particular group and the part of the Russian population they can be said to represent.

To collect sms from the informants I published a web form where they could register their messages by reading them from their phones, typing them in a text box and then submitting them to my database by the click of button. To have as much material as possible for my analysis, I asked the informants to register both messages they had sent and received. Date and time of each message was also registered. By sending each informant a link to a copy of the form, which had a unique identification number assigned, I could monitor which and how many messages every informant entered into the database. Informants were explicitly asked to enter their message exactly as they appeared on the screen of their mobile

phone, without any corrections or other modifications. (A screenshot and full translation of the web form and instructions can be found in Appendix A.) The collection of messages took place between the 19th of March and the 27th of April 2009. After the collection was completed I began structuring and analysing my database using Microsoft Excel. Four messages not containing any Russian and a few obvious duplicates were deleted. For technical reasons and since my corpus is not very large, I have not used any specific software designed for linguistic analysis. All categorisations of messages and related statistics have been produced more or less "by hand" in Excel.

There are shortcomings in the methods I have used to collect and analyse messages, but I find them acceptable with consideration to the time and resources that have been at my disposal for the study. When informants re-type messages from their mobile phones there is an obvious risk that they make conscious or unconscious corrections of for example spelling. Of course, spelling mistakes that where not in the original message can also be added by wrongful keyboard strokes. Despite the fact that anonymity was guaranteed and has been maintained throughout the whole process, informants may also feel reluctant about sharing certain kinds of messages, for example very intimate ones. In a worst-case scenario, this behaviour could cause some linguistic feature only present in a specific type of message to be absent form the corpus. The most treacherous aspect of analysing the database is clearly categorisation of single features in a message. For example, the same missing letter in a word could be regarded as a misspelling, odd transliteration or intentional reduction. In cases such as this example I have simply chosen what occurred to me, after consulting reference literature where possible, as the most likely interpretation. In some messages, understanding grammatical constructions and deciphering transliteration is hindered by a lack of knowledge about the context, which shows the advantage of alternative methods where a closer contact is maintained with the informants (see for example Hård af Segerstad, 2002, pp. 209-211).

The outcome of my online survey is a corpus of 272 sms, containing a total of 3330 words. The messages are written with both the Cyrillic and the Latin

alphabet, 158 and 114 respectively, and they vary in length from a single word up to as many as 75. Some informants reported messages that they had stored in their mobile phones and as a result my corpus displays messages sent between the 13th of January 2008 and the 25th of April 2009.

The major problem with my corpus is its size. The low number of messages makes any results and conclusions unreliable. Another problem is the uneven distribution of messages from different informants. My results are of course affected by the fact that some informants contributed more than 30 messages and other less than 10. I also have no possibility to know the origin of the incoming messages, which are a considerable part of the corpus.

What you are about to read

The rest of this thesis is divided into three main chapters, covering theoretical background, observations from my corpus and a concluding comparison and discussion respectively. To avoid a repetition of facts and introduce elements in the most logical order, the theoretical background contains facts necessary for the understanding of my results, while data used for comparison to other languages are introduced later, in the final chapter.

Background – Ocherednaia SMiSlogramma

The following pages provide background facts for my work. Working with sms, it is necessary to understand the technological background and the characteristics of the typical text message input interface. Even more important is the knowledge of what is specific for sms language and what common linguistic features it has. To be able to analyse Russian in sms one also needs a basic understanding of how the Cyrillic alphabet is used and transliterated. In addition to this, a brief discussion of some grammatical features of Russian is needed as a backdrop to my work. These four themes will be presented in the following.

Technology and sms

The technological framework surrounding sms changes rapidly. Both hardware and software for producing and handling messages have undergone a rapid evolution since text messaging had its major breakthrough in the late 1990s.

One of the main technological aspects that are significant for my study is the length of messages. Due to specific standards of encoding, an sms written with Latin letters may contain up to 160 characters, while a message containing one or more non-Latin, for example Cyrillic, letters can only consist of 70 characters (Bieswanger, 2007, p. 1). Another circumstance affecting message length is that sms are normally paid for at a fixed price per unit and not per character.

Despite the recent development towards phones with full keyboards and larger, touch-sensitive screens that make text input easier, a majority of users still produce their sms via a traditional 12-key keypad. Similarly, most messages are read on relatively small screens. Predictive text input technology makes input faster and easier, but brings new problems in the form of wrongful predictions and other technology-induced typing mistakes. Overall, typing an sms is inconvenient and "short messages over mobile phones or SMS are arguably the most distorted form [of computer mediated texting language]" (Choudhury et al, 2007, p. 158).

In addition to these factors, the asynchronicity of texting (the fact that sender and receiver do not necessarily have to be "online" at the same time) is also worth noting. In combination with the characteristics given above, this feature means "one could expect to find strategies to save time, effort and space in a corpus of SMS messages" (Hård af Segerstad, 2005, p. 35).

Language in sms – general observations

One of the most basic questions regarding sms language is its categorisation as either written or spoken communication. Ling (2005) concludes that text messaging "seems to be trans-linguistic drag queen" (sic!), displaying characteristics of both spoken and written communication, and adding some unique features. Communication via sms is like speaking in the sense that it is immediate (on a linguistic rather than technological level), personal and informal. The basic format, on the other hand, suggests that sms communication would be similar to writing and this is to some extent confirmed by looking at the relatively low occurrence rates of adjectives and adverbs among the most common words. Some of the unique and ambiguous sides of sms relate to the experience of dialogue, the common spontaneity and the perceived privacy of messaging. A more clear position for sms as mainly a spoken form communication is taken by Safi (2007), based on the observation that the two have much in common when it comes to interactivity and dependency on shared context and background. Hård af Segerstad summarises that no clear distinction can be made and that "the linguistic and communicative practices of text-messages emerge from a particular combination of technological affordances, contextual variables and interpersonal priorities" (2005, p. 35).

The social setting and situational context of sms communication is of a certain importance when dealing with the use of language in the messages. Crystal (2008, pp.105-119) draws a basic distinction between social and informational functions of sms. In the social sphere there are examples of text messages being used by the Samaritans' help centres, for Japanese speed dating and to maintain social

relationships by forwarding anecdotal chain messages. The informational uses include submitting queries to search engines, push-type advertising where information is sent to clients as it becomes available and government agencies communicating with the population. With a more statistical approach, Doering (2002) has listed the distribution of 1,000 German sms according to five text function categories suggested by Brinker's text type model:

Function	Examples	Frequency
Contact	"Have you forgotten me?!? :-(" "Goodnight my love! […]"	54%
Information	"Hi how are you. I'm a bit bored, I've been sitting in front of the TV all day, the last 4 days have been the worst." "got a computer virus, don't open any of my emails!sandra"	33%
Appeal	"Mum, I need some sliced bread from Aldi" "Could you write to web.de. I need your help."	5%
Obligation	"ok, I'll be there in half an hour. Shall I bring something along?" "I am truly sorry for my behaviour. You are much too important for me. I'll really try and change. […]"	1%
Declaration		0%
Impossible to categorise or commercial		7%

Table 1, German sms by text type (adapted from Doering, 2002, p. 2)

This listing suggests that social relations are the main concern of sms. Ling (2005) provides a slightly different categorisation. Here, a sample of 882 Norwegian sms are classified according to theme:

Theme	Examples	Frequency
Coordination	"The car is done so we can get is (sic!) at 4"	33%
Grooming	"Good that it went so well with your math exam. You are smart, Love, grandma"	17%
Answers	"I have taken care of that"	14%
Questions	"Have you caught any fish?"	11%
Information	"I found the sponge it was in the cork in the bottle"	6%
Commands or requests	"Remember to buy bread"	6%
Personal news	"We are enjoying ourselves in the sun and good weather"	5%
Other	"Thanks for the birthday present.love s"	9%

Table 2, Norwegian sms by theme (adapted from Ling, 2005, p. 338)

In conclusion, these two summaries indicate that personal relations and coordination are the main social contexts of sms.

Before moving on to the specific traits of sms language, a few words ought to be said about some general tendencies. Hård af Segerstad (2002) underlines, referring to Doering, that brevity is central among the linguistic features of sms, as the medium "both permits and forces people to express themselves concisely" (2002, p. 70). Further, "[f]eatures that are characteristic of spoken language, such as dialectal words, interjections and prosody are verbalized and spelled out in SMS" (2002, p. 70). Androutsopolous and Schmidt (2001) agree that reduction overall is distinctive for sms, and add creative language use as the second main identifier. The similarities between sms and chat communication are pointed out by Schlobinski et al (Hård af Segerstad, 2002, p. 71): the use of graphic means, the hybrid nature of sms language as both written and spoken and the varying conventions for writing. Finally, Safi states that "[i]t is assumed that SMS

syntactic and lexical choices by the texters are not so different from a child language" (2007, p. 1).

An accessible six-part introduction to the particular linguistic features of sms language is provided by Crystal (2008, pp. 37-62). Pictograms and logograms are visual symbols and letters respectively, used to represent words, parts of words or other sounds. A second category is initialisms, the practice of reducing one or a number of words to their initial letter or letters. The third common feature is omissions, where one or more letters are omitted from a word. Fourthly, Crystal brings up non-standard spellings, which can be motivated as shorter forms or sometimes markers of informality or social and regional belonging. Shortenings differ from omissions in that they remove an entire element, grammatically or semantically defined, from a word. Crystal labels his last category "genuine novelties" and basically extends it to include previously unseen exaggerations and combinations of the preceding writing techniques, as well as the parallel use of sms specific expression forms from different languages. Looking specifically at means of reducing text in sms, it is convenient to separate syntactical (related to deleting entire words and phrases) and lexical (changing and reducing within the word unit) approaches. Also, graphical means can be seen as a form of reduction. Doering (Hård af Segerstad, 2002) gives the following overview of the most common syntactical reductions in German sms:

Type of reduction	Examples
Deletion of subject (especially subject pronoun)	[Ich] Komme spatter Heim...!
Deletion of preposition, article and possessive pronoun	Weißt du was [der] Eintritt kostet
Deletion of copula-, auxiliary- or modal verbs (+XP)	[Bist du] Schon wieder zurück aus [Ø] Urlaub?
Deletion of verb and subject pronoun; Telegram style	[Hast du] Lust, dann komm vorbei?

Table 3, syntactical reductions in German sms (quoted from Hård af Segerstad, 2002, p. 201)

While Doering only lists various acronyms and abbreviations as lexical short forms, Hård af Segerstad's (2002) analysis of 1,152 Swedish sms provides a more complete summary of the linguistic features of sms:

Category	Features
Punctuation	Omitting punctuation
	Unconventional punctuation
	Omitting blank space
Spelling	Mispredictions
	Spoken-like spelling
	Split compounds
	Consonant writing
	Conventional abbreviations
	Unconventional abbreviations
	Either all capitals or all lower case
	Exchange long words for shorter
Grammar	Omission of subject pronoun
	Omission of VP (copula, auxiliary or modal verb + preposition)
	Omission of article, preposition, possessive pronoun
Graphical (non-alphabetical) means	Emoticons
	Asterisks
	Symbol replacing word

Table 4, linguistic features of Swedish sms (quoted from Hård af Segerstad, 2002, p. 215)

A somewhat comparable description, focused on non-standard orthographic forms, is put forward by Thurlow (2003). Here, the data sample is a collection of 544 British English sms:

Form type	Examples			
Shortenings	mon	'Monday'	goss	'gossip'
	tog	'together'	manch	'Manchester'
Contractions	hm	'home'	cld	'could'
	msging	'messaging'	w'end	'weekend'
G clippings	Drivin	'driving'	sortin	'sorting'
Other clippings	Ankl	'ankle'	couldn	'couldn't'
Acronyms	DI	'Detective Inspector'		
Initialisms	T now	'the'	TTFN	'Ta ta for now'
	Wbs	'write back soon'		
Letter/number homophones	no1s	'no ones'	2moro	'tomorrow'
	m8	'mate'	BCNU	'Be seeing you'
'Misspelllings' and typos	Rember	'remember'	flics	'flicks'
Non-conventional spellings	Sumtime	'sometime'	mite	'might'
	ure	'you're'	ruff	'rough'
Accent stylisation	laf	'laugh'	av	'have'
	cuz	'cause'	aught	'nothing'

Table 5, non-standard orthography in English sms (adapted from Thurlow, 2003)

Another account of novel ways of writing in sms is presented by Anis (2007). According to his findings, the unconventional spellings (defined by Anis as "neography") of sms are similar to those of the language used for chatting in French:

Feature	Examples		
Omission of accents	peut-etre		'peut-être'

Substitution of *k* for *qu*	je croyais ke	'je croyais que'
Phonetic realisations	Moua	'moi'
Truncations	comme d'hab 'comme d'habitude'	
Suppression of vowels	Tjrs	'toujours'
Syllabograms	C	'c'est'
Numerals substituting for syllables	qq1	'quelqu'un'

Table 6, non-standard orthography in French sms (adapted from Anis, 2007, p. 90)

According to Anis, these observations are noteworthy since "[m]ore strictly than for other European languages, the rules of French orthography are considered to be absolute law" (2007, p. 89). All of the above categorisations add up to a picture of language use in sms as quite similar to that of chat language but with some adaptations to the specific conditions of sms, such as brevity.

A Russian perspective on sms

The research done specifically on Russian sms is limited and varies in depth and direction, but nonetheless there are some interesting observations available. For transcription, I am using the BSI system with a few modifications for increased clarity (j for й and yo for ё), unless otherwise stated.

Sidorova states, in an appendix arguing for a more positive and less panicky view of the development of sms language, that " typological features of Russian and specific features of Russian mentality … condition significant differences (at least at this stage) between Russian sms language and English texting" (2006, p.174, translation: RH). Further, she gives an example that, compared to the view of researchers on the English language that the tendency to reduce is almighty, "the Russian lingual personality is not willing to sacrifice for the sake of compression suffixes indicating subjective valuation" (2006, p.181, translation: RH). The issues that a person composing an sms must deal with are: compression

of the text, transliteration, wordplay and "grooming" (Ru: похлопывание). Compression of the text is carried out on five main levels:

Level	Typical applications	Examples
Graphical	Shortenings	щас, (shchas) сейчас, (sejchas, 'now') skoko, (скоко) skol'ko (сколько, 'how much')
Syntactical	Deletion of personal pronoun Use of impersonal constructions (elimination of verbal phrase)	[Я] оформляю заказ (oformlyayu zakaz, 'setting up the order') Мне на частнике или на метро? (mne na chastnike ili na metro?, 'cab or subway for me?')
Morphological	Exchange of more or less equivalent verb forms	еду уезжаю (uezzhayu, 'I'm leaving') говорил поговорил (pogovoril, 'spoke')
Word invention	Clippings	универ университет (universitet, 'university')
Lexical	Use of shorter words, Russian or foreign	тут здесь (zdes', 'here') shop магазин (magazin, 'store')

Table 7, compression of text in Russian sms (adapted from Sidorova, 2006, p. 176)

Transliteration presents not only the problem of expressing Cyrillic letters with Latin ones, but also that of determining the level of preciseness in the transliteration process, as some letters can be transliterated in a more or less

"obvious" manner. Basically, the time and effort saved by the sender by using a very "reduced" transliteration will instead have to be spent by the receiver to "decode" the message.

Interestingly, Sidorova points out that "one of the most important features of sms communication is its overall positive, friendly character" (2006, p. 178, translation: RH). A majority of sms contain some form of grooming and a significant percentage of message content is phatic elements, which are rarely reduced, such as: "Hello! How are you?", "Good morning, trumpet calling!" and "Get ready, you're about to dream of me". Three particular styles, albeit with some modifications, can be identified within the genre of sms: high (official, business, academic), middle (neutral) and low (spoken language) style. According to Sidorova, the Russian idea of language has traditionally developed in a paradigm of these three styles, which carries over to the genre of sms. (Sidorova, 2006, pp. 176-182)

Another view of the characteristics of Russian sms is proposed by Kostyuchenko (2006). The most prominent features are related to vocabulary, syntax, graphics and orthography. The vocabulary of sms is influenced by spoken language and often jargon, slang, foreign words and abbreviations. Written constructions that imitate spoken sounds are also used to convey the sentiment of the sender. Russian sms syntax is characterised by the dropping of verbs, pronouns or other parts of speech. Formal subordinating means of expression are often replaced by "looser" syntactic connections. Kostyuchenko underlines that the use of non-standard graphical means, such as Latin letters, does not affect the content of messages. The basic principle for understanding the orthography of Russian sms is 'it is written as it sounds'. (Kostyuchenko, 2006, pp. 40-42)

The Cyrillic alphabet and its peculiarities

The Cyrillic alphabet has its roots in the Medieval Greek uncial script and was invented during the 9th to 10th century (Cyrillic alphabet, 2009). Letters to denote sounds that were not shared between Greek and the Slavic languages were taken

form the Glagolitic script, which had been developed earlier for the Old Church Slavonic language (Kyrillisk skrift, 2009). The original Cyrillic alphabet had 43 letters, many of which have been eliminated over time to suit the needs of different languages and dialects (Cyrillic alphabet, 2009). However, for a relatively long period of time the Cyrillic alphabet was not used to write Russian, because the language was rarely written at all. It was not until the 18th century that authors firmly established a standard for written Russian. The final scheme for writing Russian is attributed to authors Karamzin and Pushkin (Ryska. 2009).

The present day Russian Cyrillic alphabet has 33 letters. There are 20 consonants, 10 vowels, a semi-consonant/semi-vowel and two letters that do not represent sounds themselves, but have a modifying function on other letters (Wade, 2000, p. 1). One of the most basic distinctions in Russian pronunciation is that between hard and soft (palatalised; pronounced with the centre of the tongue raised to create a "j-sound") consonants. This peculiarity has a great influence on orthography. As far as the 10 vowels are concerned they are made up of five pairs, each pair consisting of a neutral and an "iotated" version of more or less the same vowel sound (the vowels are listed here with their approximate transliteration in brackets):

Neutral	а (a)	о (o)	э (e)	у (u)	ы (y)
Iotated	я (ya)	ё (yo)	е (ye)	ю (yu)	и (i)

Table 8, Russian vowels (adapted from Wikland, 1993, p. 265)

These two sets of vowels are mainly used to mark whether a preceding consonant should be pronounced as hard or soft (Wade, 2000, p. 5), a variation that is possible in 15 of the 20 consonants. Iotated vowels are also used at the beginning of words or after another vowel to indicate that the current should be preceded by a "j-sound". (**я**блоко = yabloko, 'apple' and Ма**я**ковский = Mayakovskij) In the case where a soft consonant ends a word or is succeeded by a hard consonant, the soft sign (ь) is needed to mark this softness. (мат**ь** = mat', 'mother' and реал**ь**но = real'no, 'for real') The hard sign (ъ) is used exclusively to mark a hard consonant

succeeded by an iotated vowel (in authentic Russian words this will happen only when certain word stems are modified with certain prefixes). (объяснить = ob''yasnit', 'explain') The letter й, a consonant by definition, can in practice only occur immediately after a vowel (which in turn can be either neutral or iotated), either in word final position or preceding a consonant. (чайка = chajka, 'seagull' and май = maj, 'May') (Wikland, 1993, p. 268) In addition to these conventions set by pronunciation, Russian orthography also incorporates a number of so called spelling rules. The main spelling rules are: The letter э is never written after a consonant in authentic Russian words; instead e is used, even if the consonant is hard. After ч, щ (both always soft), ж, ш (both always hard), г, к and х the letter ы is replaced by и, я by a and ю by у. The three last consonants also are never succeeded by the soft sign. Finally, o is replaced by e if it would have occurred in an unstressed position after ч, щ, ж, ш and ц (always hard). (Wade, 2000, p. 15; Wikland, 1993, p. 267)

There are many different systems for transliterating Cyrillic script to Latin. These systems reflect different needs and lingual backgrounds of those who constructed them. Logically, the main differences in transcription are found when looking at those letters that lack a natural Latin counterpart. Most prominent in this category are the vowels я and ю, as well as the "exotic" consonants ш, щ, ж, ч, ц and х. A selection of systems of transliteration is presented in appendix B.

What makes the Russian language special

In this short introduction to Russian language structure and grammar I have chosen to focus on three areas, which are of particular importance for my study: personal pronoun omission, verb omission and punctuation.

The Russian verbs conjugate according to grammatical person (first, second or third; singular or plural) in the present and future tenses (Wade, 2000, pp. 240-241). Past tense verb forms are in agreement with the subject in terms of gender and number, meaning they occur in one of the four forms: masculine, feminine, neuter or plural. The system of agreement applies to adjectives as well. This is of

relevance since they can sometimes be used to indicate for example current states of being and feelings. (я доволен, я довольна = I'm satisfied; male and female speaker respectively) Altogether, these grammatical circumstances lead to a situation where, "in spoken Russian, pronouns in general are often omitted" (Wade, 2000, p. 120). In some constructions the third person plural verb form (in sayings and informal contexts sometimes also the second person singular form) has to be used without a pronoun to indicate that the speaker is referring to people in general (Wade, 2000, p. 139). A third person pronoun ('it') is also missing from impersonal constructions of the type 'it seems to me…', since this function is carried by the verb form alone (Wikland, 1993, p. 212). Zdorenko (2008) has studied subject omission in the Russian National Corpus and draws a number of conclusions: In informal spoken Russian the subject pronoun is dropped in 29 percent of the cases where it would be possible, which is considerably higher than in any form of written language. The drop rate for second person pronouns is slightly higher than that for first person contexts. Some verb forms, most notably понимаешь (ponimaesh', 'you understand') but also знаешь (znaesh', 'you know'), have more or less grammaticalised as discourse markers when used without pronouns (Zdorenko, 2008).

Apart from the omission of personal pronouns, the Russian language structure also allows verbless constructions in some cases (no pun intended). The verb быть ('to be') does not have any conjugation for the present tense (Wade, 2000, p. 256). This means that correct Russian speech or text has numerous sentences where there is no (conjugated) verb. It is perfectly normal to hear phrases, which if translated strictly word by word would mean 'I Russian', 'On wall picture' or 'He stronger me'. When writing, a dash ('-') is often used in these situations to mark emphasis (Wade, 2000, p. 256). The impersonal 'it is' is also expressed without a verb form. This feature can be combined with a pronoun in the dative case to render fully functional sentences along the lines of 'cold for him' or 'for me boring'.

When it comes to punctuation, Wade points out that "[r]ules of punctuation are, in general, more rigorously applied in Russian than in English" (2000, p. 20).

The main differences between the languages are found in the use of the comma, the dash and in the punctuation of direct speech. Also, exclamation marks tend to be more frequent in Russian and can sometimes be used in the middle of a sentence. In contrast to English, Russian requires a comma for example between clauses joined by co-ordinating conjunctions, between a principal and a subordinate clause and to distinguish relative from main clauses. "The dash is extremely widespread in Russian" (Wade, 2000, p. 27) and apart from actng as an indicator of forms of 'to be' it can also replace other words (often verbs) that would otherwise have been repeated or can be understood from the context. The dash is further used in place of the comma to introduce unexpected turns of events or sharp contrasts. (Wade, 2000. pp. 23-28)

Results – Ну что, по пиву?))

This chapter presents the findings that have resulted from my work. The presentation begins with general facts and ends in more specific issues. When an example of an sms or an excerpt from one is given, the original text with no modifications is marked with quotation marks and followed by an approximate English translation (and where it is relevant, a transcription to Cyrillic or Latin script) in square brackets. When translating I have strived to maintain the formatting and style of the original text, rather than the exact meaning of every word.

General observations – common words, style and writing techniques

The following table gives some basic facts about my corpus, which consists of 272 messages and 3330 words in total:

Category	Average	Median
Characters per message	76,5	64
Words per message	12,6	10,5
Characters per word	4,9	5

Table 9, corpus overview

The median is included here to indicate a "truer" mean value as it less affected by extremely short or long messages than the average measurement. Further analysis of these figures is carried out in my concluding comments.

The most common words overall are presented in the table found below:

Rank	Word		Occurrences	Percentage
1	в; v, w	('in, at')	116	3,5
2	я; ya, ja, ia	('I')	101	3,0

3	не; ne ('not')	75	2,3
4	и; i ('and')	72	2,2
5	на; na ('on, at')	70	2,1
6	с; s ('with')	54	1,6
7	ты; ty, ti ('you')	50	1,5
8	а; a ('and, but, whereas')	42	1,3
9	что; 4to, chto ('that, which')	41	1,2
10	как; kak ('as, like')	30	0,9
10	у; u, y ('at, with')	30	0,9

Table 10, most common words in the corpus

The two words ranking as number ten are tied at thirty occurrences. It is noteworthy that all words on this list are functional words, which carry little or no meaning by themselves. The most common words with more content are сегодня (segodnya, 'today'), завтра (zavtra, 'tomorrow'), все (vsyo, 'everything, everyone, all') and спасибо (spasibo, 'thank you, thanks').

As there is no absolute method for judging the level of language, I will present a number of messages that summarise the corpus. I have chosen two messages to represent the extremes of casualness and formality and then added some "average" messages between them. This overview begins from the most informal:

"Ну вот и весна! Только не ощущается нифига))))))))" [So here is spring now! Just can't feel a damn thing))))))))]

"На тренировку идешь?" [Are you coming to the training?]

"Я в Калин-де, так что можно встретиться в любое время!" [I'm in Kalin-d, so we can meet at any time!]

"Хочу :) и про очиститель воздуха хочу узнать. В субботу будет возможность :)" [I want to :) and I want to find out about the air cleaner. On Saturday it'll be possible :)]

"Ок. Спасибо за помощь!" [Ok. Thanks for helping me!]

"Если взялвсе, что хотел оттуда, то бросай... Раз время пришло..." [If you tookeverything you wanted from there, then leave it… Time has come…]

"А я вечером думала,что в разговоре не до конца выразила мысли и как бы продолжала беседу..." [Well, in the evening I thought I hadn't finished expressing my thoughts in the chat and somehow continued the conversation…]

"#Name#, naberite menya posle vashego obscheniya s glavnym inzhanerom (sic!) #Name#, pozhaluista. #Name#, #Company#." [#Name#, please telephone me after your contact with the head engineer #Name#. #Name#, #Company#.]

The last message in this listing makes use of the polite second person plural to address the receiver.

When comparing the messages in my corpus to standard written Russian, I found words and expressions borrowed or adapted from spoken language (colloquialisms) in approximately 40 percent of all messages. These excerpts contain some examples:

"**Che-to** mne do nee ne dozvonitsya." [**Like**, my call doesn't go through to her. | **Чё-то** мне до неё не дозвониться.]

"**Eeee... Ty pro CD?**" [**Uhhh… You about the CD?**]

"Да, **фотки - обалдеть**." [Yeah, the **pics – awesome**.]

"Я **ваще** не готовилась" [I **tooootally** haven't prepared myself] (ваще = вообще)

"**Набери** меня - денег **нету** в телефон!" [**Buzz** me! **Outta** money on phone!]

"в ЖЖ **вообще 100 лет** не была." [haven't been on ZhZh in **like a 100 years**] (ЖЖ = Живой Журнал, Russian blog community)

The examples given here indicate both lexical and grammatical colloquialisms and some of these aspects will be treated more rigorously further on, in the section on reductions.

Another general observation is the occurrence of non-Russian language. Foreign elements occur in 10,0 percent of the messages in my corpus and there is an average of 0,19 non-Russian words per message. All cases but one ("fors mazhor" [force majeure | форс мажор]) concern English words and phrases:

"chuvstvuu sebya vinovatym, **as usual**." [I feel guilty, as usual.]

"**ok, we'll try this**..spasibo tebe za info!" [ok, we'll try this..thank you for the info!]

"#Name#, перезвони, **плз**, дело есть :)" [#Name#, call me back, plz, something has come up :)]

"**Ньюз?**" [News?] (Ньюз would be transliterated as 'Nyuz')

"sulit tekstualnii kontakt i potentsialnii **connect**." [promises literal contact and potential connect.]

As indicated by the examples, anglicisms are mainly found as distinct expressions and phrases, rather than single English words merged into a Russian context (seen in the last example).

Punctuation is another interesting issue at this general level of observation. Punctuation at the end of a sentence is left out in 16 percent of messages and in all of these it occurs only once per message. In the vast majority of the cases the full stop or another punctuation mark is left out after the last word of the message:

"Под яблонями юга москвы или это акации" [Under the southern moscow apple trees or is that acacias]

"Ya ne dozvonilsya do tebya 14-go (ermitazh, rukami), no zakinul albom na tochku s diskami. Mozhesh obmenyat na svoy, esli budesg mimo" [I couldn't get a hold of you the 14th (the ermitazh, with hands), but I dropped the album at the point with the discs. You can change it for yours, if yuo'll be nearby]

"у меня опять нет инета. если что - звони домой" [i'm out of i-net again. if anything – call home]

It can be concluded from these few brief excerpts that the use of the comma and dash are slightly irregular. Interestingly, commas are indicated in many situations where they could have been omitted without major effect on the interpretation of the message:

"Sluchaetsa. Eshe ne fakt**,** chto on budet, u nas zasedanie kafedry." [It happens. Still it's no fact**,** that he's coming, we have a faculty meeting.]

"Mne vchera skazali**,** chto #Name# iz Kyshtyma nashelsya!" [I heard yesterday**,** that #Name# from Kyshtyma is back!]

"хорошо**,** но помни о моем немецком." [good**,** but don't forget about my German]

"Ты хоть понял**,** кто я?" [Have you at least understood**,** who I am?]

"Edu v teatr**,**potom skoree vsego osyadu tam po delam do vechera." [On my way to the theatre**,**then I'll probably sit there with work till the evening.]

There is also a clear tendency to use of symbols to replace common punctuation marks, which occurs in 21 percent of messages. The smiley :-), which is often duplicated or reduced to :) or), is clearly the most common symbol:

"Не зайду **:)** передумал. Может завтра..." [Won't stop by **:)** changed my mind. Maybe tomorrow…]

"Ti mne voobwe 430 dala**)))** ya tebe 75 rub doljna teper'**)))**" [But you gave me 430**)))** now i owe you 75 roub**)))**]

"V Kaire! V starbakse s amerikankoj kofe pju **)** Ty polu4ila moju smsku?" [In Cairo! Drinking coffee with an American woman at starbucks **)** Did you get my sms?]

"Я еду мультики смотреть **:)**" [I'm on my way to watch cartoons **:)**]

"В тетради жуткий БАРДАК**:@**про голову-молчу**:S**" [My notebook is a horrible MESS**:@**won't say anything about my head**:S**]

It is not possible, however, to detect any specific tendency that would suggest in what textual contexts a symbol is more likely to be used to replace an ordinary punctuation mark.

Cyrillic vs Latin

When the corpus is split into a Cyrillic and Latin part, the following figures can be calculated:

Category	Cyrillic average	Cyrillic median	Latin average	Latin median
Characters per message	60,7	52,5	99,1	86
Words per message	10,4	9	15,8	14
Characters per word	4,6	4	5,1	5

Table 11, corpus overview by alphabet

The numbers reveal no great surprises. The Latin messages contain more words and letters presumably because they can be longer for technical reasons. The difference in word length between the two alphabets can be explained by the simple fact that some Cyrillic letters are normally transliterated with two Latin ones.

The tables listing the most common words in Cyrillic and Latin messages show some differences. It is mainly the same words on both lists, but not in the same order:

Rank	Word		Occurrences	Percentage
1	я	('I')	57	3,6
2	в	('in, at')	47	2,9
3	на	('on, at')	39	2,4
4	и	('and')	35	2,2
5	не	('not')	32	2,0
6	а	('and, but, whereas')	24	1,5
6	что	('that, which')	24	1,5
7	ты	('you')	22	1,4
8	с	('with')	20	1,3
9	привет	('hello')	15	0,9
9	у	('at, with')	15	0,9
10	мне	('to me, for me')	14	0,9

Table 12, most common words in Cyrillic messages

Rank	Word		Occurrences	Percentage
1	v, w	('in, at')	69	4,0
2	ya, ja, ia	('I')	44	2,5
3	ne	('not')	43	2,5
4	i	('and')	37	2,1
5	s	('with')	34	2,0
6	na	('on, at')	31	1,8
7	ty, ti	('you')	28	1,6
8	a	('and, but, whereas')	18	1,0
8	no	('but')	18	1,0
8	как; kak	('as, like')	18	1,0

9	4to, chto	('that, which')	17	1,0
9	menya, menia, menja	('me')	17	1,0
10	u, y	('at, with')	15	0,9

Table 13, most common words in Latin messages

Naturally, these two lists display the same tendency as table 10, which covers both alphabets. Small, functional words completely dominate over longer words with more content.

One additional "numerical" factor that is of interest is the proportion of messages that are close the maximum amount of characters allowed in a single message (70 for Cyrillic and 160 for Latin):

Alphabet and length	Number of messages	Percentage
Cyrillic 65-70	18	11,3
Cyrillic 71-76	8	5,1
Latin 155-160	9	7,9
Latin 161-166	3	2,6

Table 14, messages near maximum length for one message

This measurement should give an indication of the pressure to compress one's writing when using the different alphabets.

The most striking difference between Cyrillic and Latin messages is the omission of the soft, and hard, sign. The Cyrillic part of the corpus displays no omission of these letters, while there are 158 instances of correct usage. In contrast, the Latin part of the corpus only has 54 of a total of 164 expected soft and hard signs marked (a ratio of 32,9 percent). The following messages exemplify the use and non-use of the soft sign:

"Пока нисколько – я не успел насобира**ть** тебе крышек, мне плати**ть** нечем." [None so far – I didn't have time to gather the cups for you, I have nothing to pay with.]

"Spasibo za teplye slova i pozhelaniya. Ya rad, chto ty **dvigaeshsya dalshe** I **napolnyaesh** svou **zhizn** novymi smyslami i radostyami. U menya vse prekrasno. Vesna!" [Thanks for the warming words and wishes. I'm happy that you are **moving along** and **fulfilling** your **life** with new meanings and delights. Everything is great with me. Spring time! | Спасибо за тёплые слова и пожелания. Я рад, что ты **двигаешся далше** и **наполняеш** свою **жизн** новыми смыслами и радостями. У меня всё прекрасно. Весна!] (correct forms: двигаеш**ь**ся, дал**ь**ше, наполняеш**ь**, жизн**ь**)

"Kto menya palit? Ya **tolko** vchera **podumat** uspel!" [Who's bugging me? I **only** had time to **think** yesterday! | Кто меня палит? Я **толко** вчера **подумат** успел!] (correct forms: только, подума**ть**)

"Mne nachinaet **nehvatat' kakogo-nibud** last.fm dlya kino. Ne uspevayu **pridumivat**, chto **posmotret**." [I'm beginning to **miss some kind** of last.fm for movies. I don't have time to **decide** what to **watch**. | Мне начинает **нехватать какого-нибуд** ласт.фм для кино. Не успеваю **придумыват**, что **посмотрет**.] (correct forms: какого-нибуд**ь**, придумыва**ть**, посмотре**ть**)

Examples of soft signs being both dropped and included in the same message are relatively scarce, as only 7 out of 52 messages where one or more soft signs are omitted also contained one or more soft signs (illustrated in the forth example above).

In general reductions and compressions are more frequent in Latin messages. Filtering out all identified reductions, including abbreviations, deletion of words, omitted letters and lack of space between words and punctuation leaves 50,6 percent of the Cyrillic messages while only 28,1 percent of the Latin messages remain. This relation is equally present even if reductions where a part of a word is removed (which could be the result of transcription for the Latin messages) are not taken into account. Examples of reductions will be given in the following section.

Spelling mistakes, in the sense of involuntary exceptions from standard orthography, is another field where there is a notable difference between Cyrillic and Latin messages. Among the 17 messages containing what can be reasonably

labelled misspellings (as opposed to transcription induced oddities) only three, or 17,6 percent, are Cyrillic. The majority of the spelling mistakes is made up of single letters, either in the wrong position or completely left out.

Looking at the presence of languages other than Russian, the Latin messages are again clearly dominant. Only 13,0 percent of the messages containing foreign words are written with Cyrillic script.

To conclude this section it is also worth noting that there are no significant differences between messages written in Cyrillic and Latin respectively as far as the frequency of symbols, creative use of language (wordplay, puns) or colloquialisms is concerned.

Words, symbols, abbreviations, clippings and punctuation – means of reduction

My corpus contains 48 omitted words spread over 39 messages. This gives an average of 0,18 words omitted per message for the corpus as a whole. Verbs are most frequently left out, followed by pronouns. There is also a small group of other types of omitted word.

Starting with verbs, I have identified 20 omitted verb forms spread out over 18 messages. Further comments can be found below, after the examples of omitted verb forms:

"Davno! **Tri nedeli kak**." [Long time! **Three weeks since**.]

"Книжку на компьютере читаю, **а ты что?**" [Reading a book on the computer, **what you up to?**]

"**Я через минут 30.**" [I'll in about 30 minutes.]

"Zavtra **dumau v Etazhi** na otkrytie v 18 i v a-2 k 20." [Tomorrow **I think to Etazhi** for the opening at 18 and to a-2 around 20.]

"Mitya ne protiv v subbotu, a ya **by segodnya v zal** (esli ty), no popozhe (18-19?)" [Mitya's not against on Saturday, but I **could to gym today** (if you), but a little later (18-19?)]

"Ja kopiju sdelal, no segodnja ne uspeju uzhe otdat'. **Davaj zavtra posle 12?**" [I have made the copy, but now there's no time to give away today. **What about tomorrow after 12?**]

As these examples illustrate, most of the omitted verbs are verbs that denote departing, moving, arriving (verbs of motion) or meeting someone. In many cases a form of пойти (pojti, 'to go') could easily have been inserted into the message. Also forms of собираться (sobirat'sya, 'to intend [to go somewhere]') and встретиться (vstretit'sya, 'to meet') could be added to many of the messages without any change to the other components.

There are 17 instances in 14 messages where a pronoun would have added information to the content of the message. (Present and future tense constructions without pronouns, where the subject is defined through verb conjugation have not been included here.) This information (most often identifying the sender) is in many cases quite obvious or easily understood if the message is seen as part of a conversation:

"Привет! Уже приехал? Поздравляю в праздником (sic!), **желаю быть** если не защитником отечества, то хотя бы своих убеждений!" [Hello! Already arrived? I congratulate in for the holiday (sic!), **wish to be** if not protector of the motherland, then at least of your convictions!]

"privet! eto #name# (#organisation#). **tol'ko chto prochitali** soobshenie tvoe.." [hello! this is #name# (#organisation#). **just read** your message..]

"**Gulyala** mimo teh igrovyh avtomatov." [**Walked** past those gambling machines.]

"Привет!:)**так рад** твоей смс-ке.У меня все хорошо.Только вышел из метро." [Hello!:)**so happy** about your sms.Everything's fine with me.Just came out of the subway.]

"Видно действительно **решила** самостоятельно Россию исследовать" [Apparently **decided** to investigate Russia independently]

Naturally, almost all of the dropped pronouns can be rather securely identified as either я (ya, 'I') or ты (ty, 'you'). There are a few messages where, as in the last

example, a third person is referred to without the use of a pronoun (however, the Russian construction does establish the gender of the third person, in this case female).

The clearly left out words that are neither pronouns nor verbs amount to 11, distributed on 8 messages. Given the small number and fragmented nature of the samples, it is difficult to see any obvious tendencies:

"поздравляю! :) желаю, **чтобы все!** :)" [congratulations! :) I wish, **that everything!** :)]

"Возьми с собой **на всякий** Хауса и что нужно для зала" [Bring House and what you need for the gym **just in**]

"s chasom novoy i xoroshey muzyaki. **Kak smotrish?**" [with an hour of good and new music. **What you think?**]

"**Ya xotel** zakazat-podarit." [**I like** to order as a gift.]

The most interesting, and at the same time the most questionable, deletion occurs in the last example here. The conditional particle бы (by, no proper translation available) seems to be left out, but there is also a possibility that the sender wants to express a previous desire, which is what the phrase means in its current form.

As a result of figures being used to replace individual letters in transcription the corpus features a large amount of non-letters. Not surprisingly, figures are also commonly used to write numbers. However, in the entire corpus there are only two examples of symbols somehow replacing whole words:

"Ya segodnya v zal + sauna." [Today I'm going to the gym + sauna.]

"U tebja est spisok tem dlja diploma ili para tem kotorye ty mne mojesh porekomendovat. A to u menja **0** informacii!)))" [Do you have a list of themes for the diploma or a couple of themes, which you can recommend me. Otherwise I have **0** information!)))]

Other than the observation that the practice to replace words with symbols is virtually inexistent in my corpus, these two examples lead to no useful information. There is also one occasion where a symbol is used for a part of a word consisting of more than one letter, but this is not done with a Russian word (the case referred to is 'n8' for the English 'night').

More common than using symbols to replace words is the habit of creating initialisms, abbreviations based on the first letter of each component word. In the corpus, 20 initialisms are distributed on 18 messages:

"при покупке от 3000**р**. чудесный купальник в подарок!" [with purchase from 3000**r**. wonderful swimsuit for free!]

"а еще придется **егэ** сдавать, просто прикол!" [and on top of it I have to do my **ege**, just funny!] (егэ = единый государственный екзамен, central state exam)

"we4erom 3awtra,**ok** 23**h**. **T.4.**dawaj po obstojatel'stwam" [tomorrow evening,**ab** 23**h**. **S.t.**let's go by the circumstances]

"Привет!**НА** выходных отмечаем **др** #name#, и в саду у меня субботник.." [Hello!**IN** the weekend we celebrate #name's# **bd**, and I'll be working in the garden..]

Most of the abbreviations used in the corpus are accepted and frequent in modern Russian, such as р for рубль (rubl', 'rouble') and др for день рождения (den' rozhdeniya, 'birthday'). Many intialisms have an English origin, like сиди ('CD') and smska ('sms'). The most interesting abbreviations are t.4., which also occurs in another message in Cyrillic script as т.ч., for tak 4to and так что respectively ('so that, so then') and the two letter initialism ok for около (okolo, 'about, approximately').

Clippings, defined as one or more letters being deleted from a word, is one of the more frequent means of reduction I have encountered in the corpus. There are 34 messages that contain 36 more or less clipped words. This gives an average of 0,13 clipped words per message in the corpus as a whole. Clippings appear in many different extents and styles:

"Ty ne sobiraeshsya v tsentr fotografii na italyantsev? ili esche **kuda-** ya voobsche ne v teme, chto seychas interesnogo?" [You're not going to the photo centre for the Italians? or **smplace** else I'm totally out of the loop on what's interesting now?] (куда-нибудь reduced to куда-)

"dumaem, kuda poiti segodnja, **kakienibud'** predlozhenija mozhet byt' es't :)" [we're thinking, where to go today, maybe there are **sum** suggestions :)] (какие-нибудь reduced to какиенибудь, which is difficult to illustrate in translation)

"Privet!Segodnya pokon4u s Palanikom i **pryam** s utra voz'mus' za tvou knigu!" [Hello!Today I'll finish Palanik and **righ** from the morning begin with your book!] (прямо reduced to прям)

"Может посидим **где-н**, часиков в 9!" [Maybe we can hang out **somew**, around nineish!] (где-нибудь reduced to где-н)

"**Спс**, до завтра!" [**Thx**, see you tomorrow!] (Спасибо reduced to Спс)

"**S4az** gljanu)))" [I'll take a look **rite nw**] (Сейчас reduced to Счаз)

"Я не приду. Уже на **след** неделе." [I won't come. Only **nxt** week.] (следующей reduced to след)

As one can see from these examples, indefinite adverbs (куда-нибудь, какие-нибудь and similar) are prone to be clipped in various ways. More traditional Russian clippings, such as инфо (info, 'information') and админ (admin, 'administrator'), are also present in the corpus. The present active verbal participle следующий (sleduyushscij, 'next, following') is clipped in four different messages by four different senders, twice to следущий and twice to след. This is the single most clipped word in the corpus. The specific form of clipping known as 'consonant writing' is virtually absent in my corpus, with only three instances, which are all submitted by the same sender.

Leaving out punctuation or the spacing between a punctuation mark and the first word of the next sentence are also quite frequent means of reduction. These two techniques are applied in a comparable number of messages, 43 and 41

respectively, but there is a major difference in the amount of unique occurrences. While the leaving out of punctuation never appears more than once per message, dropping the spacing immediately after a punctuation mark occurs as many as 11 times in one message. In total, there are 0,29 reductions of the latter kind per message in the whole corpus:

"Ясно**..В**ыздоравливайте скорей**!А** то нам надо обязательно встретиться**!:)**" [Clear**..G**et well soon**!B**ecause we absolutely have to meet**!:)**]

"O Bozhe**!K**osmos!" [Oh Lord**!C**osmos!]

"Привет**.С**пасибо**:)я** вспомнила**,ч**то мы давно не встречались**.К**огда**?:)**" [Hi**.T**hanks**:)i** remembered**,t**hat we haven't met in long while**.W**hen**?:)**]

Apart from these deletions there are also examples of spacing being omitted between words in the middle of a sentence, but they are much less common (19 messages in the whole corpus show this feature) and quite often appear to be unintentional. Another special category of omitted spacing is when it appears in conjunction with the dash. This particular reduction is rather widespread, with 10 occurrences produced by 6 different senders. Typically, both spaces around the dash are omitted, but in some cases only the one following it. Looking at the issue from the reverse angle, there are 181 messages (66,5 percent), which show no irregularities in spacing or punctuation.

Writing Russian in Latin

My focus here is on systems of transliteration, meaning sets of sets of letters, letter combinations and symbols used together by senders to create their messages. The main differences appear in the treatment of iotated vowels and 'complex' consonants, which have no direct Latin equivalent.

There is a basic difference between systems that use Latin letters according to sound value and those that are based on the visual appearance of the letter. The difference is illustrated in the two following examples:

"(pisgu (sic!), chtoby ne razbudit, esli ty spish) da, u menya na russkom." [(writimg, not to wake you, if you're sleeping) yes, mine is in Russian. | (писгу, чтобы не разбудит, если ты спиш) да, у меня на русском.]

"V4era y menia predzawita bila-yspewno,da*e na nay4nyu konferenciu pozvali." [Yesterday was my predefense-successful,I was even invited to a scientific conference | Вчера у меня предзащита была-успешно,даже на научную конференцию позвали.]

These two messages show a clear difference in the treatment of the letters у (u, y), ч (ch, 4), ы (y, i) and ш (sh, w). There are also other letters that display similar variations. Hybridisations (as in the second example below) and exceptions are not uncommon, most notably with the letter x. In most cases, it is written as a Latin x by users of the sound-based system, even though it can be argued that h or kh would be a better match:

"Kak naschet povtorit uspex proshlogodnix blinov kak-nibud na nedele?" [What about somehow repeating the success of last years bliny this week? | Как насчёт повторит успех прошлогодних блинов как-нибуд на неделе?]

 "Privet! Nu 4to ect novoct" [Hello! So what any news | Привет! Ну что ест новост]

Due to these circumstances it is difficult to draw a clear line between the two systems, but in general the letters у and ш   ('u' and 'sh') appear to be the most reliable indicators of what system is preferred in a particular message.

Within the sound-based system it is possible to identify a few interesting tendencies. They mainly concern the letters я and ю. Altogether, the corpus displays five variations for the transliteration of я (ya, ja, ia, ua, a) and four for ю (yu, ju, iu, u). Basically, these variants are used as matching sets where possible. However, there is a considerable preference for 'u', which is clearly unmatched by

'a'. Also, random exceptions and mixing is not uncommon. 'ja' and 'ju' show a strong correlation, and messages with this type of transliteration are highly likely to apply 'j' for the Russian й. The same relation holds for 'ya' and 'y' for й, but it cannot be as clearly confirmed. It is worth noting in connection to this that users who prefer this system also tend to use 'y' for the Russian vowel ы. An interesting parallel between consonants and vowels is the clear relation between the use of 'c' for the Russian ц and the non-use of 'ya'. Another of the very few tangible tendencies relating to consonants is that users who choose 'zh' for ж are prone to write ш as 'sh'.

To conclude this section, I have compiled a table of all transliterations for chosen Russian letters that are found in my corpus (this table should not be seen as any indication of what transliteration is more common or which variants occur together). This compilation may also be corrupted by spelling mistakes and attempts to transliterate intentional deviations from correct Russian spelling:

в	ё	ж	й	у	х	ц	ч	ш	щ	ы	ю	я
v	e	zh	j	u	x	ts	ch	sh	sch	y	yu	ya
w	ye	z	y	y	h	c	4	w	w	i	ju	ja
	ew	j	i					sg	sh	yi	iu	ia
	o	*									u	ua
												a

Table 15, variations in transliteration

Conclusion – Экшн, тонкий юмор, всего в меру

This conclusion aims to compare my results to those of studies on other languages, and also to my background material. In addition to this, I give some independent comments.

Comparing to other languages and previous research

To compare general statistics for sms written in different languages I have compiled corpus data from a number of different sources, which explains why the comparison is inconsequent at times. The age, gender distribution and social background of the participants who contributed with their messages are not always clearly stated, but where these facts are given, they appear reasonably comparable to those of my informants.

Author	Bieswanger, 2007		Doering, 2002	Ling, 2005	Hård af Segerstad, 2002	My current corpus
Language	German	English	German	Norwegian	Swedish	Russian
Average number of characters per message	95	91	78		64	76,5
Average number of words per message			13	5,54 (men) 6,95 (women)	14,77	12,6

Table 16, length of sms in different languages

Few conclusions can be drawn from this compilation, but it is clear that Russian does not differ extremely from the other languages. The variations between languages is easily explained by the differences in language structure, for example the use of prepositions in one language (meaning many words with few characters,

as in English) contra the use of suffixes in another (meaning few words with many characters, as in Russian, and to an even greater extent Turkish).

The only listing of the most frequent words in sms that I have encountered is provided by Hård af Segerstad (2002, p.214). She has chosen to treat punctuation marks as separate words, but it is still of interest to compare her findings with mine:

Rank	Russian		Swedish	
1	в; v, w	('in, at')	.	
2	я; ya, ja, ia	('I')	!	
3	не; ne	('not')	?	
4	и; i	('and')	,	
5	на; na	('on, at')	du	('you')
6	с; s	('with')	på	('on, at')
7	ты; ty, ti	('you')	i	('in')
8	A; a	('and, but, whereas')	jag	('I')
9	что; 4to, chto ('that, which')		det	('it')
10	как; kak	('as, like')	är	('am, are, is')
	у; u, y	('at, with')		

Table 17, most common words in Russian and Swedish sms (third column adapted from Hård af Segerstad, 2002, p. 214)

Not considering the punctuation marks and two words that have no proper equivalent in Russian ('det' and 'är'), the four remaining Swedish top ten words all have a matching word among the seven most common words in my corpus. This can be seen as an, although weak, indication that Russian sms language is not exceptional in this aspect.

Still on the more general level of studying language differences in sms, Crystal remarks that "[w]hen they wish to express emotion, or be playful, Japanese

texters make more use of emoticons and other pictograms; French texters rely more on written language, abbreviated or full, and make more use of wordplay" (2008, p. 147). Again, judging from my corpus, Russian does not qualify for any of the extreme positions on this scale. There might be a tendency to prefer symbols, especially considering that they are many times duplicated, for more emotionally charged messages, but on the other hand there are enough proofs of creative language to say that Russians also apply this technique with some regularity.

It appears to me that the overall level of language in my corpus is comparable to that, which is suggested in other studies. I have chosen to include examples of Swedish, Norwegian and English sms language:

"Jo, är i gbg sen 1 vecka. men har gjort ngt idiotiskt. skaffat ett 2 veckors städjobb i hamnen. måste upp 5 på morron & är DÖD när jag kommer hem.. Ringer i kväl" [Well, been in gbg 1 week. But have done smth stupid. got myself a 2 week cleaning job in the harbour. gotta get up at 5 in the mornin & i'm DEAD when i get home.. will call to nigh]

(adapted from Hård af Segerstad, 2002, p.228)

"hei du! joda det går bare bra med mæ vettu! prøvde å ring dæ nettopp men nr var iikke i bruk? vet ikke om æ kommer opp eller ikke..." [hiya! well things are just fine with me u know! tried calling you just now but the nr was nnot in use? don't know if i passed or not…]

(adapted from Ling, 2005)

"Thanx 4 the time we've spent 2geva, its bin mint! Ur my Baby and all I want is u!xxxx"

(adapted from Bieswanger, 2007, p.6)

Judging from these examples I would consider Russian to be at the more conservative end of the scale, closer to Norwegian than to English. It is of course precarious to do this sort of comparison without access to, and sometimes even a proper understanding of, the messages that others have worked with. It is,

however, my conclusion, based on my corpus and the examples from other languages at my disposal, that Russian in sms is relatively similar to the Russian used in other informal writing contexts.

Going more into detail, Bieswanger (2002) presents a contrastive analysis of shortening strategies in English and German sms. Bieswanger uses an English corpus of messages presumably written by college students and a section of a German corpus chosen to match this group (male and female, age 17 to 30). The analysis treats six categories of shortenings: initialisms, clippings, contractions, letter-/number-homophones, phonetic spellings and word-value characters. I have chosen four of these categories of shortenings to compare with the results from my corpus (all figures refer to the number of shortenings per message in the complete corpora):

Type of shortening	English	German	Russian
Initialisms	0,03	0,13	0,07
Clippings	1,21	0,53	0,13
Letter-/number-homophones (e. g. using 2 in '2nite')	1,62	0,00	0,00
Word-value characters (e. g. using x for 'kiss')	1,24	0,10	0,01

Table 18, frequency of some shortenings in English, German and Russian sms

This comparison shows that the messages I have studied contain considerably fewer shortenings than Bieswanger's samples. The only category where my results are comparable is initialisms.

Crystal states in his chapter devoted to linguistic features of sms in languages other than English that "[t]he same set of abbreviatory processes described for English … seems to be use everywhere, with just a few variation reflecting the properties of the individual languages" (2008, p. 146). What Crystal regards as

general processes is pictograms (symbols replacing words and syllables), logograms (single letters replacing words and syllables, based their sound when pronounced), initialisms, omissions, shortenings and code-mixing with English (words, phrases and "textisms") (2008, pp.128-145). This behaviour is exemplified in more than ten languages worldwide (Crystal, 2008, p. 130). However, it is worth noting that the frequency of pictograms, homophonic abbreviations imported from English and logograms, is extremely low in my corpus of Russian sms.

Looking at the previous research on Russian in sms, my results show no dramatic contrast to the ideas put forward by Sidorova (2006) and Kostyuchenko (2006). On the contrary, my findings seem to confirm earlier observations on style, level of language, closeness to spoken language and the reluctance to abbreviate. Further, I see no tendencies that are not in line with the general Russian language structure and grammar.

Why Russian sms are special

What are then the most distinctive features of the Russian language as it is applied in sms? It is tempting to answer that question by claiming that there are no such distinctive features, but this would clearly be an exaggeration. What I wish to express is that the most striking feature in my eyes is the lack of attributes that seem unique to Russian sms language. There is, of course, the preference for Latin transliteration and the abundance of smileys, but looking beyond this, the findings that set the messages in my corpus apart from Russian as we normally know it are relatively scarce. As we have seen, it is easier to find differences when comparing to how other languages are used in text messaging. Consonant writing is practically non-existent and letters or symbols replacing syllables hardly ever occur in genuine Russian words. Whether writing with Latin or Cyrillic letters, Russians appear to stay true to their standard orthography.

I see a few possible explanations for this. Firstly, the Russian language has a strong correlation between individual letters and phonemes (the basic audible units

of speech). This means that Russian words are to a large extent pronounced "as they are written", a fact that leaves little room for replacing letter combinations with other single letters while maintaining the same pronunciation. Apart from this, Russian also has specific features, for example suffix inflection to indicate case, verb inflection, verbless impersonal constructions and the active use of the dash, which allow quite extensive syntactical reduction without compromising grammatical correctness. This in turn, could be assumed to reduce the need for lexical reductions (reductions within words). A completely different reason why "English style" texting is less common in Russia is simply that the population traditionally have had less knowledge of and exposure to Western language and culture.

However, one cannot help but to wonder if there are no more extreme examples of sms Russian. The following dialogue is featured on a Russian language training website:

- T de?	[- Whr u @?
- B 6koJIe, a 4to?	- In skool, why?
- DaBai vsTpeTuMcR! R o4 sosky4.	- Letz meet up! Im so bore.
- R To#e. Ok. 3BoHu. zzz!	- Me t00. Ok. Call. zzz!]

(adapted from Teleshkola, 2007)

Apparently, this sample is more radically reduced and creatively transliterated than any message in my corpus, but it is still relatively undistorted in comparison to many English examples. I will return to the question of where to find more extreme messaging in the next section.

In bright contrast to the seemingly prevailing strive to maintain proper orthography is the clear preference to drop all soft signs when transliterating. This is a curious tendency within the otherwise quite conservative and exact transliteration systems applied in the corpus. I believe that a major reason why the soft sign is often omitted is simply that the apostrophe symbol normally used to

denote it is rather difficult to type on a mobile phone. It is not completely unlike that some mobile phone users do not even know how to produce this symbol when entering a message. Most Russians are probably also aware of the common practice of ignoring the soft sign in transliteration unless extreme accuracy is required. In addition, this circumstance in itself means that people in general are used to interpreting "Latin Russian" without soft signs and are therefore not bothered by it in text messages.

Looking at my study against the background of sms research as whole, I think it can serve a double purpose. Hopefully, this survey can function as a reminder that many particular aspects of a language must be taken into consideration when studying its usage in a limited setting, such as sms. Secondly, I would like to think that my work could help bring attention to the idea that the world of text messaging might not be as uniform as we sometimes see it.

Final remarks

Looking back at my work, I slightly regret that I have not had the possibility to conduct it on a larger scale. Working with more informants, a larger corpus and a broader perspective would surely have been rewarding. Extending my investigation to other age and social groups would probably have resulted in more exciting findings.

Relying on material like mine, where a large amount of the content is produced by a small amount of the informants, is far from ideal. I have done my best to compensate for this situation, but there are still some problems that cannot be avoided. But after all, my ambition is to explore and to give examples, and I believe that I have reached this goal.

Except for the appealing option to generally widen my approach, for example by analysing sms sent by children and adolescents, some narrower aspects of the corpus have also caught my attention. Specifically this has to do with "inventive" and unconventional use of the language in sms. It would be interesting to conduct a thorough follow-up on the 't. 4.' abbreviation (for так что) and its frequency of

occurrence in a larger sample of texts. The possible dropping of бы in conjunction хотеть opens up another exciting perspective. This could be an indication of a beginning simplification of 'I would like to…' phrases, which could possibly become established as a convenience-motivated exception from the rule that the conjunctive requires the presence of бы.

As a finishing touch, I would like to present one of the more anecdotal, perhaps even poetical, passages I have found in my corpus, which actually might also say something about the reasons behind the limited English influence on Russian sms :)

"Девочка пяти лет отказывается от Макдональдса, на вопрос почему зло отвечает, что у нее на него аллергия." [A five-year-old girl refuses to eat at Macdonald's, when asked why, she resentfully replies that she is allergic to it.]

References

Androutsopolous, J & Schmidt, G (2001), *SMS-Kommunikation: Ethnografische Gattunsanalyse am Beispiel einer Kleingruppe*, Institut für Deutsche Sprache Mannheim, accessed 8 February 2009, <http://www.mediensprache.net/archiv/pubs/1341.pdf>.

Anis, J (2007), Neography: Unconventional spelling in French SMS text messages. In: Danet, B & Herring, S (eds), *The multilingual internet – Language, culture and communication online* (pp. 87-115), Oxford: Oxford University Press.

Bieswanger, M (2007), *2 abbrevi8 or not 2 abbrevi8: A contrastive analysis of different shortening strategies in English and German text messages*, Johann Wolfgang Goethe-University Frankfurt, accessed 26 May 2009, <http://studentorgs.utexas.edu/salsa/proceedings/2006/Bieswanger.pdf>.

Boutler, C (2009), *Transliteratsiya russkogo alfavita*, accessed 30 May 2009, <http://www.russki-mat.net/trans.htm>.

Choudhury, M, Saraf, R, Jain, V, Mukherjee, A, Sarkar, S & Basu, A (2007), 'Investigation and modeling of the structure of texting language', *International Journal on Document Analysis and Recognition*, vol. 10, nr. 3-4, pp.157-174.

Crystal, D (2008), *Txting – The gr8 db8*, Oxford: Oxford University Press.

Cyrillic alphabet (2009). In: *Encyclopædia Britannica*. accessed 30 May 2009, <http://search.eb.com.ludwig.lub.lu.se/eb/article-9028431>.

Doering, N (2002), '"Have you finished work yet? :-)" Communicative functions of text messages (SMS)', *Receiver #6. The mobile Self*, accessed 29 May 2009, <http://www.receiver.vodafone.com/06/articles/index05.html>.

Hård af Segerstad, Y (2005), Language in SMS – A socio-linguistic view. In: Harper, R, Palen, L & Taylor, A (eds), *The inside text – social, cultural and design perspectives on SMS* (pp. 33-51), Dordrecht: Springer.

Hård af Segerstad, Y (2002), *Use and adaptation of written language to the conditions of computer-mediated communication* (revised version), Department of Linguistics, Göteborg University.

Kostyuchenko, A (2006), 'Yazokovye osobennosti russkikh sms', *Vestnik MAPRYAL*, no. 53, pp. 38-43.

Kyrillisk skrift (2009). In: *Nationalencyklopedin*, accessed 30 May 2009, <http://www.ne.se.ludwig.lub.lu.se/lang/kyrillisk-skrift>.

Ling, R (2005), The socio-linguistics of SMS: An analysis of SMS use by a random sample of Norwegians. In: R, Ling & P, Pedersen (eds), *Mobile communications: Renegotiation of the social sphere* (pp. 335-349), London: Springer.

Ryska (2009). In: *Nationalencyklopedin*, accessed 30 May 2009, <http://www.ne.se.ludwig.lub.lu.se/lang/ryska>.

Safi, M S (2007), *SMS Text Analysis: Language, Gender and Current Practices*, University of Management & Technology Lahore, accessed 27 May 2009, <http://www.tesol-france.org/Documents/Colloque07/SMS%20Text%20Analysis%20Language%20Gender%20and%20Current%20Practice%20_1_.pdf>.

Sidorova, M (2006), *Internet-lingvistika: Russkiy yazuk. Mezhlichnostnoe obshchenie*, Moscow: 1989.ru.

Thurlow, C (2003), *Generation Txt? The sociolinguistics of young people's text-messaging*, University of Washington Seattle, accessed 29 May 2009, <http://extra.shu.ac.uk/daol/articles/v1/n1/a3/thurlow2002003-paper.html>.

Teleshkola (2007), *Urok 3. Aktivnye protsessy v sovremennom russkom yazyke*, accessed 24 February 2009, <http://www.internet-school.ru/demo/531230?page=6>.

Wade, T (2000), *A comprehensive Russian grammar*, Oxford: Blackwell Publishers.

Wikland, L (1993) *Modern rysk grammatik*, Stockholm: Almqvist & Wiksell.

Zdorenko, T (2008), *Subject omission in Russian*, University of Alberta Edmonton, accessed on 31 May 2009, <http://corpus.byu.edu/aacl2008/ppt/153.ppt>.

Appendix A: Web form for registration of sms

This illustration shows how the form would appear on the informant's computer screen:

Сбор СМС сообщений

Цель данного сбора – научное исследование языка СМС. Ваша анонимность гарантируется.

ІИнформант:	1
Дата сообщения:	(В формате ДД/ММ/ГГ, напр. 27/02/09)
Время сообщения:	(В формате ЧЧ:ММ, напр. 17:33)

Исходящее или входящее сообщение?
○ Исходящее
○ Входящее

Точное содержание сообщения:

[Отправить] [Очистить]

Инструкции:

Запишите, пожалуйста, как можно больше СМС сообщений, посланных/принятых Вами. Объяснение контекста не требуется. Нас интересуют сообщения любого типа, даже самые короткие. Если Вы не желаете делиться каким-либо сообщением по личным соображениям, то это не является проблемой, но чем больше сообщений Вы запишите, тем лучше будет исследование. Нам очень важно, чтобы Вы записывали сообщения в точности, как они выглядят в Вашем мобильном телефоне, т.е. с сохранением русского или латинского шрифта, сокращений, смайлов, возможных описок и пр.

Примеры:

Точное содержание сообщения:
ny ya tak "just made a guess". Vsé ok.

Точное содержание сообщения:
Ау! А тебе можно заказать посылочку? Маааааленькую такую:)

Точное содержание сообщения:
Ti,kone4n,ne smoje6 zaehat utrom v inst?

Точное содержание сообщения:
Привет. У меня все норм. ты сама то как? Морозы и вправду сильные. Но скоро обещают тепло

Collection of SMS messages

The goal of this collection is to scientifically study the language of SMS. Your anonymity is guaranteed.

Informant:

Message date: (In the format DD/MM/YY, e.g.

27/02/09)

Message time: (In the format HH:MM, e.g. 17:33)

Outgoing or in-   O Outgoing

coming message?   O Incoming

Exact content of

the message:

Submit   Clear

Instructions:

Please fill in as many as possible of your sent and received SMS messages. No explanation of the context is necessary. We are interested in all kinds of messages, even the very shortest. If You do not wish to share a message, due to some personal reasons, this is not a problem, but the more messages you register the better the study will be. It is very important that You fill in the messages exactly as they appear in Your mobile phone, i. e. maintaining Russian or English script, abbreviations, smileys, possible mistypings and such.

Examples:

Exact content of the message:
well I like "just made a guess". All ok.

Exact content of the message:
Wow! And could you order a tiny package? A smaaaall one:)

Exact content of the message:

You cant,of course,stop by uni in the morning?

Exact content of the message:

Hello. Everything is allr with me. how about yourself? It really is freezing cold.
But they promise warmth soon

When a message was submitted, the following screen would appear:

Подтверждение данных

Ваше сообщение зарегистрировано. Спасибо!

Возвратиться на формуляр

Confirmation of data

Your message has been registered. Thank you!

Return to form

Appendix B: Systems of transliteration

This table is adapted from Boutler (2009) and Wade (2000). Due to technical factors, some diacritics of the BSI and LC systems are not illustrated. No diacritics, accents or apostrophes are normally indicated when transliterating outside of linguistic contexts.

Russian Cyrillic	ISO*-9 1995	ГОСТ** 7.79 2001	BSI***	LC****	Swedish	German
а	a	a	a	a	a	a
б	b	b	b	b	b	b
в	v	v	v	v	v	w
г	g	g	g	g	g	g
д	d	d	d	d	d	d
е	e	e	e	e	je	je
ё	ë	yo	ë	ë	jo	jo
ж	ž	zh	zh	zh	zj	sch
з	z	z	z	z	z	s
и	i	i	i	i	i	i
й	j	j	ĭ	ĭ	j	j
к	k	k	k	k	k	k
л	l	l	l	l	l	l
м	m	m	m	m	m	m
н	n	n	n	n	n	n
о	o	o	o	o	o	o
п	p	p	p	p	p	p
р	r	r	r	r	r	r
с	s	s	s	s	s	s / ß
т	t	t	t	t	t	t
у	u	u	u	u	u	u

ф	f	f	f	f	f	f
х	h (x)	x	kh	kh	ch	ch
ц	c	cz / c	ts	ts	ts	z / ts
ч	č	ch	ch	ch	tj	tsch
ш	š	sh	sh	sh	sj	sch
щ	ŝ (šč)	shh	shch	shch	sjtj	schtsch
ъ	ʼʼ	ʼʼ	ʼʼ	ʼʼ		
ы	y	yʼ	y	y	y	y
ь	ʼ	ʼ	ʼ	ʼ		
э	è	è	é	ė	e	e
ю	û (ju)	yu	yu	iu	ju	ju
я	â (ja)	ya	ya	ia	ja	ja

* International Standards Organisation

** Государственный Стандарт, Russian State Standard

*** British Standards Institution

**** Library of Congress